



19 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENT- UND
MARKENAMT

12 **Offenlegungsschrift**
10 **DE 102 28 597 A 1**

51 Int. Cl.⁷:
H 04 M 11/00
H 04 L 12/56

21 Aktenzeichen: 102 28 597.7
22 Anmeldetag: 26. 6. 2002
43 Offenlegungstag: 12. 6. 2003

DE 102 28 597 A 1

66 Innere Priorität:
101 58 438. 5 29. 11. 2001

71 Anmelder:
NEC Europe Ltd., 69115 Heidelberg, DE

74 Vertreter:
Ullrich & Naumann, 69115 Heidelberg

72 Erfinder:
Molina, Maurizio, 69126 Heidelberg, DE

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gem. § 44 PatG ist gestellt

54 Verfahren zum Übertragen von zeitsynchronen Daten

57 Ein Verfahren zum Übertragen von zeitsynchronen Daten zwischen mindestens zwei Endgeräten über ein Netzwerk, vorzugsweise über das Internet, insbesondere von Sprach- und/oder Videonachrichten, wobei zwischen den Endgeräten jeweils eine Verbindung mittels eines SIP-Servers aufgebaut wird, wobei der Aufbau der Verbindung mittels des SIP-Protokolls erfolgt und wobei der SIP-Server die Verfügbarkeit von Bandbreite zur Übertragung der Daten überprüft und/oder Bandbreite im Netzwerk für die Übertragung der Daten reserviert, ist im Hinblick auf eine qualitative hochwertige Übertragung von zeitsynchronen Daten bei technischer einfacher und kostengünstiger Ausgestaltung derart ausgestaltet, dass die Reservierung der Bandbreite vom SIP-Server dahingehend optimiert wird, dass hinreichend Dienstgüte zur Übertragung gewährleistet wird und nur möglichst wenig Bandbreite ungenutzt bleibt.

E 102 28 597 A 1

[0001] Die Erfindung betrifft ein Verfahren zum Übertragen von zeitsynchronen Daten zwischen mindestens zwei Endgeräten über ein Netzwerk, vorzugsweise über das Internet, insbesondere von Sprach- und/oder Videonachrichten, wobei zwischen den Endgeräten jeweils eine Verbindung mittels mindestens eines SIP-Servers aufgebaut wird, wobei der Aufbau der Verbindung mittels des SIP-Protokolls erfolgt und wobei der SIP-Server die Verfügbarkeit von Bandbreite zur Übertragung der Daten überprüft und/oder Bandbreite im Netzwerk für die Übertragung der Daten reserviert.

[0002] Verfahren zum Übertragen von zeitsynchronen Daten wie Sprach- oder Videonachrichten, wie von Telefongesprächen oder Videokonferenzen, über Netzwerke, insbesondere über das Internet, gewinnen immer mehr an Bedeutung, da hierdurch erhebliche Kostenersparnisse bei der Datenübertragung, insbesondere beim Telefonieren, möglich sind. Für die Durchführung der Übertragung von zeitsynchronen Daten zwischen zwei Endgeräten über ein Netzwerk, insbesondere unter Verwendung des IP-Protokolls – Internet-Protokoll –, sind Mittel zur Zeichengabe beim Verbindungsaufbau und -abbau erforderlich. Das SIP-Protokoll – Session Initiation Protocol – ist eines der für diese Zwecke verwendeten Protokolle. Insbesondere ist das SIP-Protokoll eines der vielversprechendsten Protokolle für Telefondienste über das Internet.

[0003] Das SIP-Protokoll ist allerdings dahingehend problematisch, dass die eigentliche Verbindung zur Übermittlung der zeitsynchronen Daten vom SIP-Protokoll nicht unterstützt wird. Diese erfolgt durch die Kommunikation der Endgeräte, in der beispielsweise die Art der Verbindung bzw. der Datenübertragung vereinbart sowie eine Kodierung zum Senden der Daten festgelegt werden. Das SIP-Protokoll unterstützt den Aufbau der Verbindung nur insoweit, als es den Austausch der Eigenschaften der Endgeräte bezüglich der Verbindung beinhaltet. Die Übertragung der Daten findet dann meist mittels des IP-Protokolls statt, was zur Folge hat, dass es keine Garantien bezüglich der Dienstgüte – Quality of Service –, beispielsweise für die zur Verfügung gestellte Bandbreite und die Paketlaufzeiten, innerhalb einer Verbindung gibt.

[0004] Dies hat zur Folge, dass bei Sprachnachrichten beispielsweise die Sprachqualität leidet, weil die Reihenfolge der übermittelten Pakete, in welche die Sprachdaten zur Übertragung unterteilt werden, nicht gewährleistet ist, Pakete beschädigt werden und/oder verloren gehen oder nur mit Verzögerung zugestellt werden. Dies führt zum Verlust von Sprach- bzw. Übertragungsqualität.

[0005] Um dieser Problematik entgegenzutreten, überprüft der SIP-Server die Verfügbarkeit von Bandbreite zur Übertragung der Daten und/oder reserviert Bandbreite im Netzwerk für die Übertragung der Daten. Dieses Verfahren ist allerdings dahingehend problematisch, dass die Überprüfung der Bandbreite und insbesondere die Reservierung der Bandbreite im Netzwerk für die Übertragung der Daten in nicht vernachlässigbarem Umfang Kapazität des SIP-Servers beansprucht sowie bei der Reservierung zur Übermittlung der Reservierungsanfragen Bandbreite kostet. Dies verlangsamt den Verbindungsaufbau und verlangt eine hohe Verarbeitungskapazität vom SIP-Server. Wird zuviel Bandbreite für die Übertragung reserviert, so werden auch Kapazitäten im Netzwerk verschwendet, was wiederum zu höheren Kosten sowie einer unnötig hohen Auslastung des Netzwerks und damit zu Einschränkungen bei der Datenübertragung führt.

[0006] Der vorliegenden Erfindung liegt daher die Auf-

gabe zugrunde, ein Verfahren zum Übertragen von zeitsynchronen Daten der eingangs genannten Art anzugeben, bei dem eine qualitativ hochwertige Übertragung von zeitsynchronen Daten bei technisch einfacher und kostengünstiger Ausgestaltung ermöglicht ist.

[0007] Erfindungsgemäß wird die voranstehende Aufgabe durch ein Verfahren zum Übertragen von zeitsynchronen Daten mit den Merkmalen des Patentanspruchs 1 gelöst. Danach ist das in Rede stehende Verfahren zum Übertragen von zeitsynchronen Daten derart ausgestaltet und weitergebildet, dass die Reservierung der Bandbreite vom SIP-Server dahingehend optimiert wird, dass hinreichend Dienstgüte zur Übertragung gewährleistet wird und nur möglichst wenig Bandbreite ungenutzt bleibt.

[0008] In erfindungsgemäßer Weise ist erkannt worden, dass in technisch besonders einfacher Weise eine optimierte Reservierung der Bandbreite dadurch erreicht werden kann, dass der SIP-Server zusätzlich zu der Überprüfung der Verfügbarkeit von Bandbreite und der Reservierung der Bandbreite im Netzwerk eine Optimierung der Reservierung der Bandbreite in der Art erfolgen muss, dass zum einen hinreichend Dienstgüte zur Übertragung gewährleistet ist und gleichzeitig nur wenig Bandbreite ungenutzt bleibt. Die Auslastung des gesamten Netzwerks wird somit optimiert, so dass für die Übertragung von Daten eine hinreichende Dienstgüte erreicht wird, ohne dass unnötige Kosten dadurch entstehen, dass Bandbreite reserviert wird, die nicht zur Übertragung von Daten genutzt wird.

[0009] In besonders vorteilhafter Weise könnte die Überprüfung und/oder die Reservierung der Bandbreite durch einen in den SIP-Server integrierten Quality Agent vorgenommen werden. Ein solcher Quality Agent könnte im Rahmen einer besonders kostengünstigen Ausgestaltung auch nachträglich in einen Standard-SIP-Server integriert werden, so dass keine neuen SIP-Server angeschafft werden müssten, sondern die bereits bestehende SIP-Server lediglich aufgerüstet werden müssen.

[0010] In weiter vorteilhafter Weise könnten die Reservierungen für verschiedene Verbindungen, vorzugsweise vom SIP-Server, zu mindestens einem Reservierungsbündel zusammengefasst werden. In diesem Reservierungsbündel wären dann die Bandbreite für eine bestimmte Anzahl von Verbindungen zusammengefasst, was zur Folge hat, dass lediglich die Bandbreite des Reservierungsbündels überprüft werden muss und nicht die jeder einzelnen Verbindung.

[0011] Im Hinblick auf eine besonders effektive Optimierung könnten in den Reservierungsbündeln Reservierungen für Verbindungen zwischen je zwei Endpunkten, vorzugsweise zwischen Access- oder Edge-Routern, zusammengefasst werden. Hierdurch könnte in besonders vorteilhafter Weise eine Skalierbarkeit des Netzwerks erreicht werden. Bei den sogenannten Access- oder Edge-Routern könnte es sich um IP-Router handeln, durch die eine große Anzahl von Endgeräten an das Netzwerk bzw. das Internet angeschlossen werden können.

[0012] Ferner könnte die Reservierung der Bandbreite für die einzelnen Reservierungsbündel erfolgen. Auch hierdurch wäre in vorteilhafter Weise eine Verringerung von Rechenleistung und Datenübertragungen erfolgt.

[0013] Im Rahmen einer besonders einfachen Ausgestaltung könnte die Reservierung der Bandbreite im Netzwerk außerhalb des SIP-Servers, vorzugsweise mittels eines Dienstgütemanagementsystems, zu welchem die Reservierungen gesendet werden, erfolgen. Das Dienstgütemanagementsystem könnte dann diese Reservierungsanfragen weiterverarbeiten. Bei diesem Dienstgütemanagementsystem könnte es sich um einen sogenannten Bandbreiten-Broker handeln, der auch Dienstgüterserver genannt wird. Nur das

Dienstgütemanagementsystem müsste dann die komplexe Aufgabe übernehmen, die Reservierung von Bandbreite entsprechend an die einzelnen Geräte im Netzwerk weiterzugeben. Weder die Endgeräte noch der SIP-Server wären darin involviert. Das Gleiche gilt für andere Netzwerkanwendungen mit anderen Anforderungen an die Dienstgüte, die ebenfalls den Bandbreiten-Broker in Anspruch nehmen können.

[0014] Da es sich bei dem Dienstgütemanagementsystem um eine gemeinsam genutzte Ressource im Netzwerk handelt und unter Berücksichtigung der Tatsache, dass in einem Netzwerk, insbesondere im Internet, die Anzahl von Verbindungen sehr groß sein kann, muss es grundsätzlich vermieden werden, dass der SIP-Server auf das Dienstgütemanagementsystem bei jedem Aufbau einer Verbindung zugreift. Der SIP-Server könnte daher Bandbreite für die einzelnen Reservierungsbündel im Voraus reservieren. Die Reservierungsbündel würden also hinsichtlich der zugeteilten Bandbreite vom Dienstgütemanagementsystem kontrolliert, aber der SIP-Server, insbesondere der Quality Agent, würde das Dienstgütemanagementsystem in der Dimensionierung und Änderung der Dimensionierung der Reservierungsbündel steuern.

[0015] Hinsichtlich einer besonders effektiven Ausgestaltung könnte die Anzahl der Reservierungen und/oder der Reservierungsänderungen beim Dienstgütemanagementsystem für mindestens ein Reservierungsbündel vom SIP-Server kleiner als eine maximale Rate R_{\max} gehalten werden. Hierdurch wären die an das Dienstgütemanagementsystem gestellten Anfragen begrenzt und ein überhöhter Zugriff auf das Dienstgütemanagementsystem und die damit verbundenen Nachteile könnten somit wirksam vermieden werden.

[0016] Zusätzlich oder alternativ könnte die Anzahl der Reservierungen und/oder der Reservierungsänderungen beim Dienstgütemanagementsystem für alle von einem SIP-Server im Netzwerk des Dienstgütemanagementsystems erstellten Reservierungsbündel vom SIP-Server kleiner als eine maximale Rate R_{\max} gehalten werden. In besonders vorteilhafter Weise könnte die Anzahl der Reservierungen und/oder der Reservierungsänderungen beim Dienstgütemanagementsystem für alle im Netzwerk des Dienstgütemanagementsystems erstellten Reservierungsbündel von allen im Netzwerk befindlichen SIP-Servern kleiner als eine maximale Rate R_m gehalten werden. Hierdurch wäre es möglich, die Inanspruchnahme des Dienstgütemanagementsystems insgesamt zu optimieren.

[0017] Im Hinblick auf eine besonders gute Optimierung könnten die mittleren Eingangsraten von Verbindungen $\lambda_1, \lambda_2, \dots, \lambda_n$ für alle Reservierungsbündel ermittelt werden. Zusätzlich oder alternativ könnten die mittleren Anzahlen N_1, N_2, \dots, N_n der in den Reservierungsbündeln zusammengefassten Verbindungen ermittelt werden.

[0018] Im Rahmen der Optimierung könnte ferner die mittlere Bandbreite b , die durch eine Verbindung durchschnittlich belegt wird, ermittelt werden. Zusätzlich oder alternativ könnte eine Wahrscheinlichkeit P_{Block} , dass die Bandbreite eines Reservierungsbündels nicht vergrößerbar ist, ermittelt werden. Somit könnte eine Wahrscheinlichkeit dafür angegeben werden, dass zukünftige Verbindungsanfragen möglicherweise abgewiesen werden.

[0019] Ferner könnte die mittlere Dauer einer Verbindung μ gemäß

$$\mu = \frac{\sum_{i=1}^n N_i}{\sum_{j=1}^n \lambda_j}$$

ermittelt werden.

[0020] Zu Beginn des Verfahrens, beispielsweise bei der Implementierung, könnten zunächst die mittleren Eingangsraten $\lambda_1, \lambda_2, \dots, \lambda_n$ und/oder die mittleren Anzahlen bestehender Verbindungen N_1, N_2, \dots, N_n und/oder die mittlere Bandbreite b geschätzt werden. Eine besonders einfache Implementierung des erfindungsgemäßen Verfahrens wäre somit möglich.

[0021] Im Hinblick auf eine besonders gute Optimierung könnten die mittleren Eingangsraten $\lambda_1, \lambda_2, \dots, \lambda_n$ und/oder die mittleren Anzahlen bestehender Verbindungen N_1, N_2, \dots, N_n und/oder die mittlere Bandbreite b regelmäßig, vorzugsweise basierend auf den tatsächlich eingehenden Verbindungen, aktualisiert werden. In besonders vorteilhafter Weise könnten die maximale Rate R_{\max} von Reservierungen und/oder Reservierungsänderungen und/oder die Wahrscheinlichkeit P_{Block} konstant gehalten werden. Dadurch wäre es möglich, eine gleich gute Optimierung der Übertragung über einen längeren Zeitraum hinweg zu erreichen.

[0022] Nunmehr könnte für mindestens ein Reservierungsbündel j die maximale Größe der Bandbreite des Reservierungsbündels $N_{j\max}$ und/oder eine Menge von Schwellwerten $T_1^j, T_2^j, \dots, T_K^j$ berechnet werden. Im Rahmen der Optimierung könnte ferner die Größe der reservierten Bandbreite für ein Reservierungsbündel j gleich dem nächst größeren Schwellwert $T_1^j, T_2^j, \dots, T_K^j$ gehalten werden. Wenn z. B. für einen Reservierungsbündel die Schwellwerte $3b, 7b, 15b, 20b, \dots$ betragen und die aktuell benutzte Bandbreite $8b$ beträgt, dann müsste die reservierte Bandbreite des Reservierungsbündels $15b$ betragen.

[0023] Die Schwellwerte werden für jedes Reservierungsbündel basierend auf einem Verfahren berechnet, das in ATM-Netzwerkumgebungen angewendet wird. Dieses Verfahren benutzt exakte Formeln, um einige Zwischenwerte zu bestimmen, und wird durch das erfindungsgemäße Verfahren durch einen zweiten heuristischen Schritt ergänzt, mittels dem die erforderliche Menge von Schwellwerten für jedes Reservierungsbündel genau bestimmt wird. Das erfindungsgemäße Verfahren umfasst demnach zwei Aspekte, nämlich die Schritte zur Bestimmung der zu reservierenden Bandbreiten der Reservierungsbündel von Verbindungen anzuwenden, ebenso wie alle Modifikationen anzuwenden, die nötig sind, um das Verfahren in einer IP-Umgebung zur Unterstützung von Dienstgüte für mittels eines SIP-Servers aufgebauten Verbindungen anwenden zu können. Der heuristische Schritt würde hierbei das bereits bekannte Verfahren zur Berechnung der Schwellwerte vervollständigen.

[0024] Im Rahmen einer besonders flexiblen Ausgestaltung könnte der nächst höhere Schwellwert $T_2^j, T_3^j, \dots, T_{k+1}^j$ gewählt werden, wenn die aktuell benutzte Bandbreite des Reservierungsbündels den bisherigen Schwellwert $T_1^j, T_2^j, \dots, T_K^j$ überschreitet. Wenn die aktuell benutzte Bandbreite des obigen Beispiels nun den Schwellwert $15b$ überschreitet, dann müsste die reservierte Bandbreite des Reservierungsbündels auf $20b$ erhöht werden.

[0025] Zusätzlich oder alternativ könnte die Größe der Bandbreite des Reservierungsbündels j gleich der maximalen Bandbreite N_j des Reservierungsbündels gehalten werden. Dies könnte dann von besonderem Vorteil sein, wenn die durch die Verbindungen benutzte Bandbreite gleich N_j ist.

[0026] Im Hinblick auf eine gute Optimierung könnten die Schwellwerte $T_1^j, T_2^j, \dots, T_K^j$, vorzugsweise für alle Reservierungsbündel, neu berechnet werden, wenn eine vorgegebene Abweichung zwischen den geschätzten und den tatsächlichen mittleren Eingangsraten $\lambda_1, \lambda_2, \dots, \lambda_n$ und/oder der mittleren Anzahl N_1, N_2, \dots, N_n von Verbindungen und/oder der mittleren Bandbreite b überschritten wird. Die weiteren Anfragen des SIP-Servers zur Reservierungsänderung

an das Dienstgütemanagementsystem würden dann auf den neu errechneten Werten basieren. Diese Maßnahme würde es erlauben, die Optimierung der Bandbreite auch im Falle eines stark variierenden Verkehrsaufkommens, beispielsweise in täglichen oder wöchentlichen Zyklen, durchzuführen.

[0027] Im Rahmen einer sehr flexiblen Ausgestaltung könnte zusätzlich zu der durch den Schwellwert $T_1^j, T_2^j, \dots, T_k^j$ definierten maximalen reservierten Bandbreite eine Reservebandbreite B_{spare} reserviert werden. Die Reservebandbreite B_{spare} würde es ermöglichen, Reserven für weitere Verbindungen zu haben, die vor einer Erhöhung der reservierten Bandbreite erfolgen, so dass die Wahrscheinlichkeit der Zurückweisung von Verbindungsanfragen verringert wird. Bei Anfragen zur Reservierungsänderung, also wenn ein Schwellwert überschritten wird, wird demnach die neue Bandbreite nicht genau auf den nächst größeren Schwellwert gesetzt, welcher der benötigten Bandbreite am nächsten ist, sondern auf eine Bandbreite, die der von dem nächst größeren Schwellwert zuzüglich des Betrags B_{spare} definierten Bandbreite entspricht.

[0028] Im Hinblick auf eine Optimierung der Rate der Reservierungen und/oder Reservierungsänderungen könnte eine Änderung der Reservierung der Bandbreite nach einer bereits erfolgten Änderung in entgegengesetzter Richtung erst dann erfolgen, wenn der Schwellwert T_k um einen Wert δ über- oder unterschritten wird. Mit einem solchen Hystereseverfahren, könnte also eine Reservierungsänderung nicht sofort wieder ausgelöst werden, wenn ein Schwellwert T_k in umgekehrter Richtung wieder über- oder unterschritten wird. Vielmehr muss ein Schwellwert $T_k - \delta$ über- bzw. unterschritten werden, bevor eine Reservierungsänderung der Bandbreite vor dem ersten Über- bzw. Unterschreiten wiederhergestellt wird. Der Wert δ könnte dabei positiv im Fall einer ursprünglich Überschreitung und negativ im Fall einer ursprünglichen Unterschreitung gewählt werden.

[0029] Schließlich sei darauf hingewiesen, dass es mit dem erfindungsgemäßen Verfahren ermöglicht ist, den Vorgang der Wartung und Rekonfiguration einer logischen Netzwerkinfrastruktur (Verbindungsbündel im Internet) für den Transport von Daten effizient und skalierbar einzurichten. Das erfindungsgemäße Verfahren ist effizient, weil es die Bestimmung der optimalen Bandbreite von Verbindungsbündeln beinhaltet, insbesondere bei gegebener aktueller Belegung der Reservierungsbündel mit Verbindungen und bei verfügbaren aktuellen Schätzungen der Last (Verbindung pro Sekunde) für ein solches Reservierungsbündel.

[0030] Ferner ist das erfindungsgemäße Verfahren skalierbar, weil die Gesamtfrequenz der Änderung der Bandbreiten für Reservierungsbündel (die Summe der Frequenzen der Bandbreitenänderung aller Reservierungsbündel im Netzwerk) unterhalb einer gegebenen Rate R_{max} gehalten werden kann, welche die maximale Gesamtrate von Anfragen ist, welche ein Dienstgütemanagementsystem von einem oder mehreren SIP-Servern annehmen kann. Eine optimale Lösung und somit die optimale Netzwerkausnutzung kann somit abgestimmt werden auf die Leistungsfähigkeit des Dienstgütemanagementsystems.

[0031] Das erfindungsgemäße Verfahren basiert auf einigen Hypothesen betreffend des verwendeten Verkehrsmodells, welche zureffend und realistisch für den sprachgebundenen Telefonverkehr sind. Ferner kann das erfindungsgemäße Verfahren mit erweiterten SIP-Servern implementiert werden. In Anbetracht der Tatsache, dass bereits ein Dienstgütemanagementsystem – Bandbreiten-Broker – existiert, würde die Produktkombination SIP-Server + Erweiterung + Bandbreiten-Broker für Kunden eine vollständige Lösung zur effizienten und automatischen Kontrolle von

Netzwerkressourcen für dienstgütesensitive Anwendungen wie den Telefonverkehr im Internet erreichen.

Patentansprüche

1. Verfahren zum Übertragen von zeitsynchronen Daten zwischen mindestens zwei Endgeräten über ein Netzwerk, vorzugsweise über das Internet, insbesondere von Sprach- und/oder Videonachrichten, wobei zwischen den Endgeräten jeweils eine Verbindung mittels mindestens eines SIP-Servers aufgebaut wird, wobei der Aufbau der Verbindung mittels des SIP-Protokolls erfolgt und wobei der SIP-Server die Verfügbarkeit von Bandbreite zur Übertragung der Daten überprüft und/oder Bandbreite im Netzwerk für die Übertragung der Daten reserviert, **dadurch gekennzeichnet**, dass die Reservierung der Bandbreite vom SIP-Server dahingehend optimiert wird, dass hinreichend Dienstgüte zur Übertragung gewährleistet wird und nur möglichst wenig Bandbreite ungenutzt bleibt.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die Überprüfung und/oder die Reservierung der Bandbreite durch einen in den SIP-Server integrierten Quality Agent vorgenommen wird.
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, dass die Reservierungen für verschiedene Verbindungen, vorzugsweise vom SIP-Server, zu mindestens einem Reservierungsbündel zusammengefasst werden.
4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, dass in den Reservierungsbündeln Reservierungen für Verbindungen zwischen je zwei Endpunkten, vorzugsweise zwischen Access- oder Edge-Routern, zusammengefasst werden.
5. Verfahren nach Anspruch 3 oder 4, dadurch gekennzeichnet, dass die Reservierung der Bandbreite für die einzelnen Reservierungsbündel erfolgt.
6. Verfahren nach einem der Ansprüche 3 bis 5, dadurch gekennzeichnet, dass die Reservierung der Bandbreite im Netzwerk außerhalb des SIP-Servers, vorzugsweise mittels eines Dienstgütemanagementsystems, zu welchem die Reservierungen gesendet werden, erfolgt.
7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, dass die Anzahl der Reservierungen und/oder der Reservierungsänderungen beim Dienstgütemanagementsystem für mindestens ein Reservierungsbündel vom SIP-Server kleiner als eine maximale Rate R_{max} gehalten wird.
8. Verfahren nach Anspruch 6 oder 7, dadurch gekennzeichnet, dass die Anzahl der Reservierungen und/oder der Reservierungsänderungen beim Dienstgütemanagementsystem für alle von einem SIP-Server im Netzwerk des Dienstgütemanagementsystems erstellten Reservierungsbündel vom SIP-Server kleiner als eine maximale Rate R_{max} gehalten wird.
9. Verfahren nach einem der Ansprüche 6 bis 8, dadurch gekennzeichnet, dass die Anzahl der Reservierungen und/oder der Reservierungsänderungen beim Dienstgütemanagementsystem für alle im Netzwerk des Dienstgütemanagementsystems erstellten Reservierungsbündel von allen im Netzwerk befindlichen SIP-Servern kleiner als eine maximale Rate R_{max} gehalten wird.
10. Verfahren nach einem der Ansprüche 3 bis 9, dadurch gekennzeichnet, dass die mittleren Eingangsdaten von Verbindungen $\lambda_1, \lambda_2, \dots, \lambda_n$ für alle Reservierungsbündel ermittelt werden.

11. Verfahren nach einem der Ansprüche 3 bis 10, dadurch gekennzeichnet, dass jeweils die mittleren Anzahlen N_1, N_2, \dots, N_n der in den Reservierungsbündeln zusammengefassten Verbindungen ermittelt werden.
12. Verfahren nach einem der Ansprüche 3 bis 11, dadurch gekennzeichnet, dass die mittlere Bandbreite b , die durch eine Verbindung durchschnittlich belegt wird, ermittelt wird.
13. Verfahren nach einem der Ansprüche 3 bis 12, dadurch gekennzeichnet, dass eine Wahrscheinlichkeit P_{Block} , dass die Bandbreite eines Reservierungsbündels nicht vergrößerbar ist, ermittelt wird.
14. Verfahren nach Anspruch 9 und 10 und ggf. einem der Ansprüche 11 bis 13, dadurch gekennzeichnet, dass die mittlere Dauer einer Verbindung μ gemäß

$$\mu = \frac{\sum_{i=1}^n N_i}{\sum_{j=1}^n \lambda_j}$$

20

ermittelt wird.

15. Verfahren nach einem der Ansprüche 10 bis 14, dadurch gekennzeichnet, dass zunächst die mittleren Eingangsraten $\lambda_1, \lambda_2, \dots, \lambda_n$ und/oder die mittleren Anzahlen bestehender Verbindungen N_1, N_2, \dots, N_n und/oder die mittlere Bandbreite b geschätzt werden.
16. Verfahren nach einem der Ansprüche 10 bis 15, dadurch gekennzeichnet, dass die mittleren Eingangsraten $\lambda_1, \lambda_2, \dots, \lambda_n$ und/oder die mittleren Anzahlen N_1, N_2, \dots, N_n von Verbindungen und/oder die mittlere Bandbreite b regelmäßig, vorzugsweise basierend auf den tatsächlich eingehenden Verbindungen, aktualisiert werden.
17. Verfahren nach einem der Ansprüche 7 bis 16, dadurch gekennzeichnet, dass die maximale Rate R_{max} von Reservierungen und/oder Reservierungsänderungen und/oder die Wahrscheinlichkeit P_{Block} konstant gehalten werden.
18. Verfahren nach einem der Ansprüche 3 bis 17, dadurch gekennzeichnet, dass für mindestens ein Reservierungsbündel j die maximale Größe der Bandbreite des Reservierungsbündels $N_{j\text{max}}$ und/oder eine Menge von Schwellwerten $T_1^j, T_2^j, \dots, T_K^j$ berechnet wird/ werden.
19. Verfahren nach Anspruch 18, dadurch gekennzeichnet, dass die Größe der reservierten Bandbreite für ein Reservierungsbündel j gleich dem nächst größeren Schwellwert $T_1^j, T_2^j, \dots, T_K^j$ gehalten wird.
20. Verfahren nach Anspruch 19, dadurch gekennzeichnet, dass der nächst höhere Schwellwert $T_1^j, T_2^j, \dots, T_K^j$ gewählt wird, wenn die aktuell benutzte Bandbreite des Reservierungsbündels den bisherigen Schwellwert $T_1^j, T_2^j, \dots, T_K^j$ überschreitet.
21. Verfahren nach Anspruch 19, dadurch gekennzeichnet, dass die Größe der reservierten Bandbreite des Reservierungsbündels j gleich der maximalen Bandbreite N_j des Reservierungsbündels gehalten wird.
22. Verfahren nach einem der Ansprüche 18 bis 21, dadurch gekennzeichnet, dass die Schwellwerte $T_1^j, T_2^j, \dots, T_K^j$, vorzugsweise für alle Reservierungsbündel, neu berechnet werden, wenn eine vorgegebene Abweichung zwischen den geschätzten und den tatsächlichen mittleren Eingangsraten $\lambda_1, \lambda_2, \dots, \lambda_n$ und/oder der mittleren Anzahl N_1, N_2, \dots, N_n von Verbindungen und/oder der mittleren Bandbreite b überschritten wird.
23. Verfahren nach einem der Ansprüche 18 bis 22,

dadurch gekennzeichnet, dass zusätzlich zum der durch den Schwellwert $T_1^j, T_2^j, \dots, T_K^j$ definierten maximal reservierten Bandbreite eine Reservebandbreite B_{Spare} reserviert wird.

24. Verfahren nach einem der Ansprüche 18 bis 23, dadurch gekennzeichnet, dass eine Änderung der Reservierung der Bandbreite nach einer bereits erfolgten Änderung in entgegengesetzter Richtung erst dann erfolgen kann, wenn der Schwellwert T_K um einen Betrag δ über- oder unterschritten wird.

- Leerseite -

Method for transmitting time-synchronous data

A method for transmitting time-synchronous data between at least two terminal devices via a network, preferably via the Internet, in particular for transmitting voice and/or video data, whereby a connection is set up in each case between the terminal devices by means of an SIP server, whereby the connection is set up by means of the SIP protocol, and whereby the SIP server checks the availability of bandwidth for the transmission of the data and/or reserves bandwidth in the network for the transmission of the data, is designed with a view to high-quality transmission of time-synchronous data in a technically simple and low-cost design in such a way that the reservation of the bandwidth by the SIP server is optimised so that adequate quality of service is guaranteed for the transmission and only the smallest possible amount of bandwidth remains unused.

Description

[0001] The invention relates to a method for transmitting time-synchronous data between at least two terminal devices via a network, preferably via the Internet, in particular for transmitting voice and/or video data, whereby a connection is set up in each case between the terminal devices by means of an SIP server, whereby the connection is set up by means of the SIP protocol, and whereby the SIP server checks the availability of bandwidth for the transmission of the data and/or reserves bandwidth in the network for the transmission of the data.

[0002] Methods for transmitting time-synchronous data, such as voice or video data, e.g. from telephone calls or videoconferences, via networks, in particular the Internet, are becoming increasingly important, given that substantial cost savings can be made in the data transmission, particularly in telephony. To carry out the transmission of time-synchronous data between two terminal devices via a network, in particular using the IP protocol (Internet Protocol), signalling means are required

during the connection set-up and cleardown. The SIP protocol (Session Initiation Protocol) is one of the protocols used for these purposes. In particular, the SIP protocol is one of the most promising protocols for telephony services via the Internet.

[0003] However, the SIP protocol causes problems in that the actual connection for the transfer of time-synchronous data is not supported by the SIP protocol. This is carried out through the communication of the terminal devices in which, for example, the type of the connection or the data transmission is agreed and a coding is defined for the data transmission. The SIP protocol supports the connection set-up only insofar as it entails the exchange of the characteristics of the terminal devices in relation to the connection. The transmission of the data then takes place mainly by means of the IP protocol, with the result that there is no guarantee in terms of the quality of service, for example in terms of the bandwidth made available and the packet transit times within a connection.

[0004] As a result, for example in the case of voice data, voice quality suffers, because the sequence of transmitted packets into which the voice data are subdivided for transmission is not guaranteed, packets are damaged and/or lost, or are delivered only with a delay. This results in loss of voice or transmission quality.

[0005] To overcome these problems, the SIP server checks the availability of bandwidth for the transmission of the data and/or reserves bandwidth in the network for the transmission of the data. However, this method caused problems in that the bandwidth check and, in particular, the reservation of the bandwidth in the network for the transmission of the data take up a significant amount of the capacity of the SIP server and use up bandwidth in the reservation for transferring the reservation requests. This slows down the connection set-up and requires a high processing capacity from the SIP server. If too much bandwidth is reserved for the transmission, capacities are also wasted in the network, which in turn results in higher costs and an unnecessarily high utilisation of the network and therefore restrictions in the data transmission.

[0006] The object of the present invention is therefore to indicate a method for transmitting time-synchronous data of the aforementioned type, in which high-quality transmission of time-synchronous data is enabled in a technically simple and low-cost design.

[0007] According to the invention, the aforementioned object is achieved by a method for transmitting time-synchronous data with the features of patent claim 1. The method concerned for transmitting time-synchronous data is designed and further developed accordingly in such a way that the reservation of the bandwidth by the SIP server is optimised so that adequate quality of service is guaranteed for the transmission and only the smallest possible amount of bandwidth remains unused.

[0008] It has been recognized in an inventive manner that an optimised reservation of the bandwidth can be achieved in a technically particularly simple manner in that the SIP server, in addition to the bandwidth availability check and the reservation of the bandwidth in the network, an optimisation of the reservation of the bandwidth must take place¹ in such a way that, on the one hand, adequate quality of service is guaranteed for the transmission and, at the same time, only a small amount of bandwidth remains unused. The utilisation of the entire network is thus optimised, so that an adequate quality of service is achieved for the transmission without incurring unnecessary costs by reserving bandwidth which is not used for the transmission of data.

[0009] In a particularly advantageous manner, the check and/or reservation of the bandwidth could be carried out by a Quality Agent integrated into the SIP server. In a particularly low-cost design, a Quality Agent of this type could also be subsequently integrated into a standard SIP server, so that no new SIP servers would have to be purchased, but rather the existing SIP server would merely need to be upgraded.

[0010] In a further advantageous manner, the reservations for various connections, preferably from the SIP server, could be combined into at least one reservation

¹ **Translator's Note:** This is a literal translation of the original German sentence, which is grammatically confused.

group. In this reservation group, the bandwidth would then be combined for a specific number of connections, with the result that only the bandwidth of the reservation group needs to be checked, and not the bandwidth of each individual connection.

[0011] With a view to a particularly effective optimisation, reservations for connections between two endpoints, preferably between access or edge routers, could be combined in the reservation groups. Scalability of the network could thereby be achieved in a particularly advantageous manner. The access or edge routers could be IP routers, by means of which a large number of terminal devices could be connected to the network or the Internet.

[0012] Furthermore, the bandwidth could be reserved for the individual reservation groups. This would also result in a reduction in computing power and data transmissions.

[0013] In a particularly simple design, the reservation of the bandwidth in the network could take place outside the SIP server, preferably by means of a quality of service management system, to which the reservations are sent. The quality of service management system could then further process these reservation requests. This quality of service management system could be a bandwidth broker, which is also referred to as a quality of service server. Only the quality of service management system would then have to perform the complex task of forwarding the reservation of bandwidth accordingly to the individual devices in the network. Neither the terminal devices nor the SIP server would be involved therein. The same applies to different network applications with different quality of service requirements, which could similarly use the bandwidth broker.

[0014] Since the quality of service management system is a jointly used resource in the network, and mindful of the fact the number of connections may be very high in a network, in particular on the Internet, it must essentially be avoided that the SIP server accesses the quality of service management system on each connection set-up. The SIP server could therefore reserve bandwidth in advance for the individual

reservation groups. The reservation groups would therefore be monitored by the quality of service management system in terms of the allocated bandwidth, but the SIP server, in particular the Quality Agent, would control the quality of service management system in dimensioning and in modifying the dimensioning of the reservation groups.

[0015] With a view to a particularly effective design, the number of reservations and/or reservation changes could be kept below a maximum rate R_{\max} by the SIP server in the quality of service management system for at least one reservation group. The requests submitted to the quality of service management system would thereby be limited, and excessive access to the quality of service management system and the associated disadvantages could thus be effectively avoided.

[0016] Additionally or alternatively, the number of reservations and/or reservation changes in the quality of service management system could be kept below a maximum rate R_{\max} by the SIP server for all reservation groups created by the SIP server in the network of the quality of service management system. In a particularly advantageous manner, the number of reservations and/or reservation changes could be kept below a maximum rate R_m by all SIP servers located in the network in the quality of service management system for all reservation groups created in the network of the quality of service management system. This would enable optimisation of the use of the quality of service management system as a whole.

[0017] With a view to a particularly effective optimisation, the mean input rates of connections $\lambda_1, \lambda_2, \dots, \lambda_n$ could be determined for all reservation groups. Additionally or alternatively, the mean numbers N_1, N_2, \dots, N_n of the connections combined in the reservation groups could be determined.

[0018] Within the optimisation, the mean bandwidth b which is occupied on average by a connection could be determined. Additionally or alternatively, a probability P_{Block} that the bandwidth of a reservation group cannot be increased could be determined. A probability that future connection requests will possibly be rejected could thus be specified.

[0019] Furthermore, the mean duration of a connection μ could be determined according to

$$\mu = \frac{\sum_{i=1}^n N_i}{\sum_{i=1}^n \lambda_i}$$

[0020] At the beginning of the method, for example in the implementation, the mean input rates $\lambda_1, \lambda_2, \dots, \lambda_n$ and/or the mean numbers of existing connections N_1, N_2, \dots, N_n and/or the mean bandwidth b could first be estimated. A particularly simple implementation of the method according to the invention would thus be possible.

[0021] With a view to a particularly effective optimisation, the mean input rates $\lambda_1, \lambda_2, \dots, \lambda_n$ and/or the mean numbers of existing connections N_1, N_2, \dots, N_n and/or the mean bandwidth b could be regularly updated, preferably on the basis of the actual incoming connections. In a particularly advantageous manner, the maximum rate R_{\max} of reservations and/or reservation changes and/or the probability P_{Block} could be kept constant. This would make it possible to achieve an equally effective optimisation of the transmission over a lengthy period.

[0022] The maximum size of the bandwidth of the reservation group $N_{j\max}$ and/or a set of threshold values $T_1^j, T_2^j, \dots, T_k^j$ could then be calculated for at least one reservation group j . Furthermore, in the optimisation, the size of the reserved bandwidth for a reservation group j could be kept equal to the next highest threshold $T_1^j, T_2^j, \dots, T_k^j$. If, for example, the threshold values for one reservation group are $3b, 7b, 15b, 20b, \dots$ and the currently used bandwidth is $8b$, the reserved bandwidth of the reservation group would have to be $15b$.

[0023] The threshold values are calculated for each reservation group on the basis of a method which is used in ATM network environments. This method uses exact formulae to define a number of intermediate values and is supplemented in the method according to the invention by a second heuristic step, by means of which the required set of threshold values is precisely defined for each reservation group. The

method according to the invention accordingly comprises two aspects, i.e. applying the steps for defining the bandwidths of the reservation group of connections to be reserved, and also applying all modifications which are required in order to be able to apply the method in an IP environment to support quality of service for connections set up by means of an SIP server. The heuristic step would complement the already known method for calculating the threshold values.

[0024] In a particularly flexible design, the next highest threshold value $T_2^j, T_3^j, \dots, T_{k+1}^j$ could be selected if the currently used bandwidth of the reservation group exceeds the previous threshold $T_1^j, T_2^j, \dots, T_k^j$. If the currently used bandwidth in the above example then exceeds the threshold value 15b, the reserved bandwidth of the reservation group would have to be increased to 20b.

[0025] Additionally or alternatively, the size of the reserved bandwidth of the reservation group j could be kept equal to the maximum bandwidth N_j of the reservation group. This could be particularly advantageous if the bandwidth used by the connections is equal to N_j .

[0026] With a view to effective optimisation, the threshold values $T_1^j, T_2^j, \dots, T_k^j$ could preferably be recalculated for all reservation groups if a predefined difference between the estimated and the actual mean input rates $\lambda_1, \lambda_2, \dots, \lambda_n$ and/or the mean number N_1, N_2, \dots, N_n of connections and/or the mean bandwidth b is exceeded. The further requests of the SIP server for the reservation change to the quality of service management system would then be based on the recalculated values. This measure would enable the optimisation of the bandwidth to be carried out even in the event of significantly varying traffic volumes, for example in daily or weekly cycles.

[0027] In a very flexible design, a reserve bandwidth B_{Spare} could be reserved in addition to the maximum reserved bandwidth defined by the threshold value $T_1^j, T_2^j, \dots, T_k^j$. The reserve bandwidth B_{Spare} would provide reserves for further connections which are established before an increase in the reserved bandwidth, so that the probability of rejection of connection requests is reduced. In the event of requests for a reservation change, i.e. if a threshold value is exceeded, the new

bandwidth is set accordingly, not precisely to the next highest threshold value which is nearest to the required bandwidth, but to a bandwidth which corresponds to the bandwidth defined by the next highest threshold value plus the amount B_{Spare} .

[0028] With a view to optimising the rate of reservations and/or reservation changes, a change could be made to the reservation of the bandwidth following a change already made in the opposite direction only if the threshold value T_k is exceeded or understepped by a value δ . With a hysteresis method of this type, a reserve change could not therefore again be initiated immediately if a threshold value T_k is again exceeded or understepped in the opposite direction. Instead, a threshold value $T_k - \delta$ must be exceeded or understepped before a reservation change of the bandwidth is restored before the first exceeding or understepping. The value δ could be selected as positive in the case of an original exceeding and negative in the case of an original understepping.

[0029] Finally, it must be pointed out that the method according to the invention enables the process of maintenance and reconfiguration of a logical network infrastructure (connection group on the Internet) to be set up efficiently and in a scalable manner for the transport of data. The method according to the invention is efficient because it entails the definition of the optimum bandwidth of connection groups, in particular for a given current utilisation of the reserve groups with connections and with available current estimations of the load (connections per second) for a reservation group of this type.

[0030] Furthermore, the method according to the invention is scalable because the overall frequency of the change of bandwidths for reservation groups (the sum of the frequencies of the bandwidth change of all reservation groups in the network) can be kept below a given rate R_{max} , which is the maximum total rate of requests which a quality of service management system can accept from one or more SIP servers. An optimum solution, and therefore the optimum network utilisation, can thus be aligned with the performance capability of the quality of service management system.

[0031] The method according to the invention is based on a number of hypotheses relating to the traffic model used, which are relevant and realistic for voice-based telephony traffic. Furthermore, the method according to the invention can be implemented with expanded SIP servers. In view of the fact that a quality of service management system - bandwidth broker - already exists, the product combination SIP server + expansion + bandwidth broker would provide customers with a complete solution for efficient and automatic monitoring of network resources for quality-of-service-sensitive applications such as telephony traffic on the Internet.

Patent claims

1. Method for transmitting time-synchronous data between at least two terminal devices via a network, preferably via the Internet, in particular for transmitting voice and/or video data, whereby a connection is set up in each case between the terminal devices by means of at least one SIP server, whereby the connection is set up by means of the SIP protocol, and whereby the SIP server checks the availability of bandwidth for the transmission of the data and/or reserves bandwidth in the network for the transmission of the data, **characterised in that** the reservation of the bandwidth by the SIP server is optimised so that adequate quality of service is guaranteed for the transmission and only the smallest possible amount of bandwidth remains unused.
2. Method according to claim 1, characterised in that the check and/or reservation of the bandwidth is carried out by a Quality Agent integrated into the SIP server.
3. Method according to claim 1 or 2, characterised in that the reservations for various connections, preferably from the SIP server, is combined into at least one reservation group.
4. Method according to claim 3, characterised in that reservations for connections between two endpoints, preferably between access or edge routers, are combined in the reservation groups.
5. Method according to claim 3 or 4, characterised in that the bandwidth is reserved for the individual reservation groups.
6. Method according to one of claims 3 to 5, characterised in that the bandwidth is reserved in the network outside the SIP server, preferably by means of a quality of service management system, to which the reservations are sent.
7. Method according to claim 6, characterised in that the number of reservations and/or the reservation changes are kept below a maximum rate R_{\max} by the SIP

server in the quality of service management system for at least one reservation group.

8. Method according to claim 6 or 7, characterised in that the number of reservations and/or reservation changes in the quality of service management system is kept below a maximum rate R_{\max} by the SIP server for all reservation groups created by an SIP server in the network of the quality of service management system.

9. Method according to one of claims 6 to 8, characterised in that the number of reservations and/or reservation changes is kept below a maximum rate R_m by all SIP servers located in the network in the quality of service management system for all reservation groups created in the network of the quality of service management system.

10. Method according to one of claims 3 to 9, characterised in that the mean input rates of connections $\lambda_1, \lambda_2, \dots, \lambda_n$ are determined for all reservation groups.

11. Method according to one of claims 3 to 10, characterised in that the mean numbers N_1, N_2, \dots, N_n of the connections combined in the reservation groups are determined.

12. Method according to one of claims 3 to 11, characterised in that the mean bandwidth b which is occupied on average by a connection is determined.

13. Method according to one of claims 3 to 12, characterised in that a probability P_{Block} that the bandwidth of a reservation group cannot be increased is determined.

14. Method according to claims 9 and 10 and, where appropriate, one of claims 11 to 13, characterised in that the mean duration of a connection μ is determined according to

$$\mu = \frac{\sum_{i=1}^n N_i}{\sum_{j=1}^n \lambda_j}$$

15. Method according to one of claims 10 to 14, characterised in that the mean input rates $\lambda_1, \lambda_2, \dots, \lambda_n$ and/or the mean numbers of existing connections N_1, N_2, \dots, N_n and/or the mean bandwidth b are first estimated.

16. Method according to one of claims 10 to 15, characterised in that the mean input rates $\lambda_1, \lambda_2, \dots, \lambda_n$ and/or the mean numbers of existing connections N_1, N_2, \dots, N_n and/or the mean bandwidth b are regularly updated, preferably on the basis of the actual incoming connections.

17. Method according to one of claims 7 to 16, characterised in that the maximum rate R_{\max} of reservations and/or reservation changes and/or the probability P_{Block} are kept constant.

18. Method according to one of claims 3 to 17, characterised in that maximum size of the bandwidth of the reservation group $N_{j\max}$ and/or a set of threshold values $T_1^j, T_2^j, \dots, T_k^j$ is/are calculated for at least one reservation group j .

19. Method according to claim 18, characterised in that the size of the reserved bandwidth for a reservation group j is kept equal to the next highest threshold $T_1^j, T_2^j, \dots, T_k^j$.

20. Method according to claim 19, characterised in that the next highest threshold value $T_1^j, T_2^j, \dots, T_k^j$ is selected if the currently used bandwidth of the reservation group exceeds the previous threshold $T_1^j, T_2^j, \dots, T_k^j$.

21. Method according to claim 19, characterised in that the size of the reserved bandwidth of the reservation group j is kept equal to the maximum bandwidth N_j of the reservation group.

22. Method according to one of claims 18 to 21, characterised in that the threshold values $T_1^j, T_2^j, \dots, T_k^j$ are preferably recalculated for all reservation groups if a

predefined difference between the estimated and the actual mean input rates $\lambda_1, \lambda_2, \dots, \lambda_n$ and/or the mean number N_1, N_2, \dots, N_n of connections and/or the mean bandwidth b is exceeded.

23. Method according to one of claims 18 to 22, characterised in that a reserve bandwidth B_{spare} is reserved in addition to the maximum reserved bandwidth defined by the threshold value $T_1^j, T_2^j, \dots, T_k^j$.

24. Method according to one of claims 18 to 23, characterised in that a change can be made to the reservation of the bandwidth following a change already made in the opposite direction only if the threshold value T_k is exceeded or understepped by a value δ .

² **Translator's Note:** In paragraph 24 above, this value is stated as " $T_2^j, T_3^j, \dots, T_{k+1}^j$ "